Journal of Engg. Research **Vol. 5 No. (3) September 2017 pp.** *50-72*

مراجعة ومقارنة طرق عرض متعددة الأبعاد لإعادة الإعمار ثلاثية الأبعاد

توشر جادهاف*، كولبير سينغ و أديتيا أبهينكار*****

* باحث، جامعة ثابار، باتيالا، البنجاب، الهند وقسم الالكترونيات والاتصالات السلكية واللاسلكية، معهد فيشوكارما لتكنولوجيا المعلومات، بيون، الهند

** الالكترونيات والاتصالات قسم الهندسة، جامعة ثابار، باتيالا، الهند

*** قسم التكنولوجيا، سافيتريباي جامعة فول بيون، بيون، الهند

* مراسلة المؤلف: *tushar.jadhav@viit.ac.in*

الخلاصة

إعادة الإعمار ثلاثي الأبعاد الأساسية من وجهات نظر متعددة تم دراسته جيدا، ولكن تكمن صعوبة المشكلة في مجال رؤية الكمبيوتر. هناك مجموعة كبيرة ومتنوعة من النهج المتاحة في السابق. تستخدم الأساليب تمثيلات مختلفة للداخل سواء كان مشهد أو كائن وقد توفر أنواعا مختلفة من المخرجات. بعض الأساليب تمثل المشهد بأكمله مثل (voxel-set) حيث يستعمل البعض مجموعات مستويات أو تمثيل شبكة مضلع. وقد يكون الناتج إما حجم أو سطح يمثل الكائن / المشهد الذي أعيد بناؤه. تعمل بعض الطرق في مساحة الصورة حيث تعمل بعض الطرق في مساحة الكائن. وقد تم تطوير هذه الطرق لتقديم حل وسط جيد بين سرعة الحساب والتعقيد الحسابي والدقة جنبا إلى جنب مع الجدوى في التنفيذ. ويتوقف اختيار طريقة معينة على متطلبات التطبيق وتوافر الموارد المطلوبة. وعلى الرغم من أن الاستعراضات السابقة المتوفرة في المراجع، فإن التقدم السريع في هذا المجال يتطلب آخر المستجدات. تقدم الورقة مراجعة ومقارنة لأحدث طرق عرض متعددة الأبعاد لإعادة الإعمار الثلاثي الأبعاد جنبا إلى جنب مع المعلومات حول مجموعات البيانات القياسية وصناديق الأدوات / البرمجيات مفتوحة المصدر. وهذا سوف يساعد الباحثين على فهم حالة آخر ما توصل اليه العلم في هذا المجال.

A review and comparison of multi-view 3D reconstruction methods

Tushar Jadhav*, Kulbir Singh** and Aditya Abhyankar***

*Research scholar, Thapar University, Patiala, Punjab, India and Department of Electronics & Telecommunication, Vishwakarma Institute of Information Technology, Pune, India

**Electronics & Communication Engineering Department, Thapar University, Patiala, India,

***Department of Technology, Savitribai Phule Pune University, Pune, India,

*Corresponding author: tushar.jadhav@viit.ac.in

ABSTRACT

Three Dimensional reconstruction from multiple views is well studied, fundamental, yet a challenging problem in the field of computer vision. There is a large variety of approaches available in the literature. The methods use different representations for input scene/object and may provide different kinds of outputs. Some methods model entire scene as voxel-set where as some use level sets or polygon mesh representation. Output may be either volume or surface representing the reconstructed object/scene. Some methods work in image space where as some methods work in object space. These methods are developed to offer a good compromise between computation speed, computation complexity and accuracy along with feasibility in implementation. Selection of a particular method depends on the requirements of application and availability of required resources. However, earlier reviews are available in the literature, fast advances in this field demand latest review. The paper presents a review and comparison of latest multi-view 3D reconstruction methods along with the information about standard datasets and tool boxes/open source software. This will help researchers to understand state of the art in this field.

INTRODUCTION

Good quality three-dimensional (3D) reconstruction of a scene or an object is a fundamental and challenging problem in the field of computer vision. The task of recovering 3D scene from two-dimensional (2D) views/images is called 2D to 3D reconstruction. Figure 1 presents the schematic for 3D reconstruction pipeline. The reconstruction pipeline comprises forward and

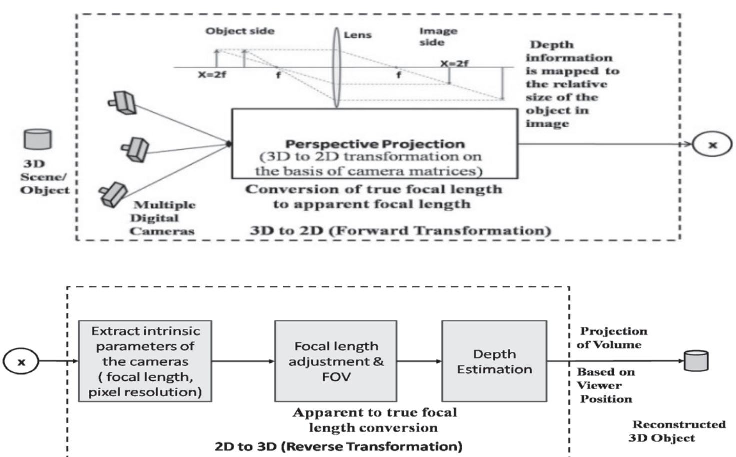

Figure 1. 3D reconstruction pipeline: forward (3D-2D) and reverse (2D-3D) transformation.

reverse transformation to convert 3D information into 2D information and vice versa, respectively. Since physical depth of the object from cameras/viewpoints is related to the true focal length, intrinsic and extrinsic parameters of cameras need to be obtained during reverse transformation. 3D reconstruction has a variety of applications such as 3D terrain rendering (Wu *et al*., 2015), virtual reality, robot navigation, augmented reality tasks, games, animations, and motion pictures. In the medical field, it is used in minimal invasive surgical techniques, 3D rendering of patients anatomy, computer guided surgeries, preoperative planning, and so on.

A large variety of 3D reconstruction approaches are available in the literature. These approaches intend to provide a good compromise between computation speed, complexity, and accuracy along with the feasibility in implementation. Different methods use different scene representations and provide different outputs (Seitz *et al*., 2006). The output may be either the volume or the surface representing a reconstructed object. Excellent reviews focusing on the earlier reconstruction methods are available in the literature (Seitz *et al*., 2006; Dyer, 2001; Slabaugh *et al*., 2001). However, the advances in this field such as reduction of computational complexity and computation time, real time implementations, large scale reconstruction, and reconstruction of the scene with high amount of details demand the latest review. This paper presents a review of multi-view 3D reconstruction methods. The paper presents a comparison between these methods on the basis of the key parameters. The reconstruction methods using same datasets and testing protocol are also compared and presented. The paper also presents the information about widely used toolboxes and datasets in multi-view reconstruction. This will help researchers to understand the state of the art in this field.

The rest of the paper is organized as follows. The next section presents the classification of multi-view 3D reconstruction methods based on the approach used for the reconstruction. These approaches are discussed in detail in different subsections. Each subsection also sheds a light on the pros and cons of the respective approach. The results of volumetric 3D reconstruction method developed by the authors are also presented in this section. The section entitled "comparison of reconstruction methods" presents a comparison of these methods on the basis of important parameters in the form of tables. This section also presents the comparison of performance of different methods using the same datasets and testing protocol. The information about widely used toolboxes and datasets is also provided in tabular form in this section. The last section presents the discussion and conclusion of the review. Although the development of multi-view 3D reconstruction methods has experienced tremendous growth in recent years, these methods need improvement in many aspects. Based on these aspects, the section gives direction to the researchers in the field of multi-view 3D reconstruction.

STATE OF THE ART

2D to 3D reconstruction methods are classified on the basis of the approach used for reconstruction. This classification is presented in Figure 2. The methods developed by the researchers using these approaches are discussed in detail in the following subsections.

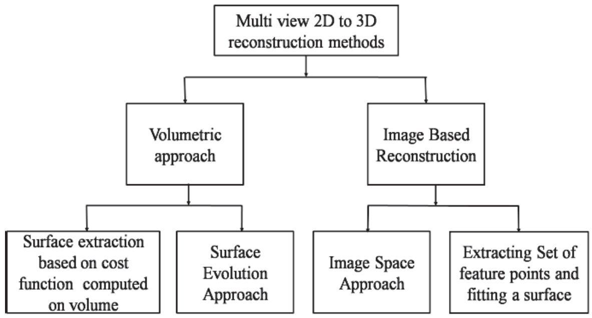

Figure 2. Classification of multi-view 3D reconstruction methods.

Surface extraction based on cost function computed on volume

The methods of this class make use of the volume of visual hull inferred from geometric intersection of the regions obtained by back-projecting each silhouette from a corresponding viewpoint (Bottino & Laurentini, 2000; Franco & Boyer, 2009). First, they compute a cost function on the 3D volume and then extract a surface from this volume. Voxel colouring methods make a single sweep through the volume and compute the colour consistency metric to reconstruct the colour consistent voxels. Seitz & Dyer introduced the voxel colour consistency constraint, in order to distinguish the surface points from other points in the scene (Slabaugh *et al*., 2001). As shown in Figure 3, when two cameras view a non-surface point, they see dissimilar

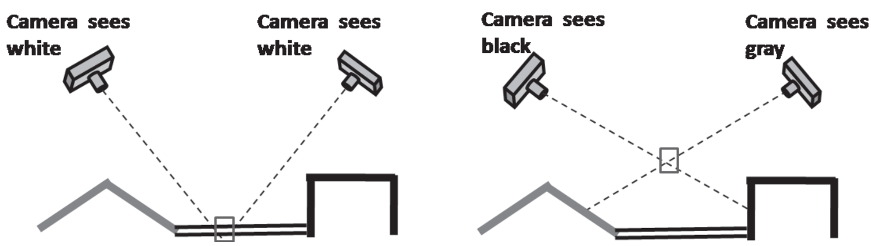

Figure 3. The left figure shows that two cameras see a consistent colour since the point lies on the surface. The right figure shows colour inconsistency for non-surface point (Slabaugh *et al*., 2001).

colours. Therefore, the views are not consistent in colour, whereas both cameras see the same colour if the point lies on the surface. The reconstruction process in these methods begins with the object volume represented by the opaque voxels. Each voxel is back-projected on the image from which the voxel is visible. The voxels are checked later for photo-consistency or colour consistency using some cost function. The voxels, which exceed a certain threshold, are carved away. The process continues until all the voxels are consistent in colour (Kimura *et al*., 1999). The method presented in the work of Adipranata & Soo (2007) uses mean charts for defining the colour consistency check. Zao & Xiao (2005) implemented the voxel colouring algorithm in HLS colour space instead of RGB colour space in order to overcome the limitation of the ordinary voxel colouring algorithm. This approach requires background-foreground segmentation. Visual hull computation is sensitive to wrong classification of the pixels in the segmentation. This happens because of the occlusion. Slembrouck *et al*. (2014) developed a self-learning algorithm for determination of occlusion of the voxels. The need of background subtraction is eliminated by defining the photo-consistency measure based on the optical sensor inputs (Moretto *et al*., 2003).

It is possible to allow the user to control the complexity of different surface regions interactively (Brisc, 2004). Such methods also work on uncalibrated images. In case of the calibrated images, the calibration data can be effectively used to determine parameterization of a voxel space and calculate the 3D polyhedron automatically further to reconstruct the volume (Feldmann *et al*., 2010). In practical situations like observing an aircraft or asteroid, the surface needs to be extracted from the silhouettes without knowing relative positions of the viewpoints (Bottino *et al*., 2000). Polygon meshes are widely used to represent the surfaces. These methods use signed octrees to determine the polygonal mesh (Montenegro *et al*., 2006). The method based on polyhedral visual hull ensures the topological properties like manifoldness and also achieves high modeling speed (Franco & Boyer, 2009). Some of the reconstruction methods construct local surfaces using a convex hull with marching cubes and combine those surfaces to complete the reconstruction (Shin & Tjahjadi, 2008), whereas some methods make use of arbitrary viewpoints to obtain viewpoint dependent image based representations to construct the visual hull. However, these methods do not provide a complete 3D model (Matusik *et al*., 2000).

Minimum cut/maximum flow algorithms on the graphs are widely used for energy minimization in computer vision. The literature presents many min-cut/max-flow methods with different polynomial time complexity (Boykov & Kolmogorov, 2004). The method based on the minimum cut of the weighted graph uses an implicit volumetric representation based on voxel occupancy to take the advantage of volumetric and image space approach (Vogiatzis *et al*., 2007). A similar graph cut algorithm, which integrates the foreground colour information and the silhouette information, is used for colour consistency field optimization (Tran & Davis, 2006). Segmentation of the object and background can be automated by identifying the object with pose of the camera instead of bounding rectangles (Campbell *et al*., 2011). The constraint based on predetermined locations, through which the object surface passes, is used to improve the performance of the method. These methods use pair-wise Markov random fields (MRF) framework based on a probabilistic approach to reconstruct the surface by computing graph cuts (Ulusoy *et al*., 2016). Real time implementation of voxel based reconstruction methods is a challenging task. Perez *et al*. (2012) developed the improved visual hull algorithm for graphics processing unit (GPU) based implementation. Improved visual hull algorithm reduces the consumption of resources.

The volumetric reconstruction methods do not need to find correspondences between different views. These methods effectively overcome the occlusion problem and also handle textureless and shiny objects. The methods are capable of reconstructing the scenes with complex geometry by using colour consistency measures. In case of the view synthesis, the use of voxel based reconstruction method is always a good choice. However, these methods demand more memory. Hence, the memory requirement is an important design consideration in volumetric reconstruction approaches. Voxel colouring based volumetric reconstruction methods do not support arbitrary placement of cameras because of the visibility constraint. Background removal is also an essential step in colour consistency based volumetric reconstruction methods.

SURFACE EVOLUTION APPROACH

Surface evolution approach is based on the evolution of the surface iteratively, by minimizing the defined cost function. The methods using this approach represent the surface using voxels, level sets, or surface meshes (Seitz *et al*., 2006). These methods mainly work in the object space. The surface evolution process starts with an initial surface and then moves with some speed along its normal. The level set theory uses partial differential equations (PDEs) to characterize such motion for the surface evolution process (Slabaugh *et al*., 2001). The convex sets based on the convex function can be defined using the level set theory, to integrate silhouette and stereo information (Kolev & Cremers, 2008). This makes the method suitable for modeling the concavities and the reconstruction of shiny metal objects. The reconstruction process is made independent of the initialization and the surface orientation by imposing silhouette constraints (Cremers & Kolev, 2011). The method based on convex optimization using volumetric labeling can reconstruct large scale scenes (Blaha *et al*., 2016). The method consumes less memory as compared to other methods for large scale reconstructions.

Surfaces are represented as deformable 3D models by using two types of representations, explicit surfaces and implicit surfaces. The use of explicit surface representations like polygon mesh makes the deformation process simpler. However, these representations suffer from problems like self-interaction and changes in topology in the form of merging and splitting. On the other hand, implicit representations make the surface evolution process easier (Ilic & Fua, 2006). The mesh evolution method proposed in the work of Zaharescu *et al*. (2011) uses an implicit surface representation and handles the topological changes robustly. Some methods initialize the surface evolution process by defining a triangular mesh followed by PDE based mesh optimization (Nghiem *et al*., 2010). The method based on multi-video data computes a temporal coherence for optimizing photo consistency and minimizes the energy function defined for hyper-surface to obtain the 3D scene (Goldluecke & Magno, 2004). Methods that initialize their surface evolution process by obtaining surface patches as depth maps apply the mesh fusion technique to construct the desired surface (Dainese *et al*., 2005). The variational approach can be combined with the level set approach to regularize the surface evolution process (Lhuillier & Quan, 2005). Reconstruction of the large scale scenes with higher amount of details is possible by using minimum s-t cut based optimization (Vu *et al*., 2009). The optimization technique is used to obtain a visibility consistent mesh from the dense point cloud.

Defining a cost/error function is one of the important steps in the surface evolution process. Weighted radial basis functions (RBFs) are used as the cost functions to generate smooth and seamless surface models from the sparse, noisy, non-uniform, and low resolution range data (Dinh *et al*., 2002). The cost function defined on the basis of discrepancy between surface normals and normal fields (Chang *et al*., 2007) and the cost function defined using specularity constraint based on surface normal (Nehab *et al*., 2008) are some other examples of such cost functions. The use of planar mesh parameterization technique eliminates the need of merging the surface normal maps (Park *et al*., 2013). The integrated surface often suffers from dent artifacts produced by depth discontinuities in the multi-view range images. The surface mesh model can be used to overcome such discontinuities (Ju *et al*., 2009). The method performing a visual hull intersection in the image plane creates the 3D structure by stacking occupancy grids on the top of each other (Khan *et al*., 2007). Some methods segment the volume into free and occupied space and extract a surface as a boundary in between them (Savinov *et al*., 2015). Kostrikov *et al*. (2014) proposed a probabilistic approach for assigning a labeling cost to each voxel. This makes the method more robust.

Space carving approaches fall under this class of 3D reconstruction methods (Seitz *et al*., 2006). Figure 4 presents the results of an efficient and simple method for volumetric reconstruction of real objects using multiple cameras. The method proposed by the authors extracts volume of the object from its multiple views acquired using pre-calibrated cameras. After extracting the silhouettes from calibrated images, the volumetric intersection is performed to obtain visual hull. Inconsistent voxels are carved away by enforcing photo consistency measures. Further, the surface extraction is done using isosurfaces. Figure 4 shows the sample views, silhouettes, and 3D reconstruction of cube and prism shaped objects. The results show that the method efficiently extracts the volume of the object. It is observed that the estimated volume of the object is approximately equal to its actual volume. The major advantage of the proposed method is its simplicity and less complexity in the implementation. The use of multiple cameras improves the reconstruction quality.

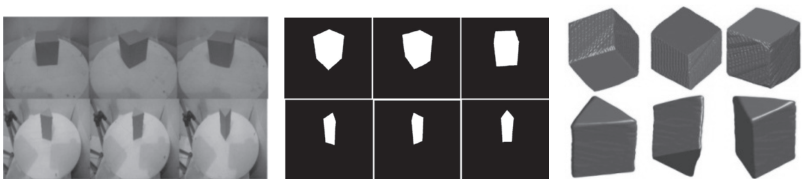

Figure 4. Sample views, silhouettes, and reconstruction of objects, cube and prism.

Voxel based reconstruction methods produce the output in the form of voxels, which is the most commonly used computer graphics format. Hence, it will be an easier task for computers to process the output further. However, the quality of 3D reconstruction is highly dependent upon the accuracy of camera calibration and the segmentation of objects in the captured views. These methods need to obtain a bounding box of the 3D scene. Hence, they are more suitable for compact scenes. These methods evolve the surface iteratively from initial guess. Hence, accuracy of such methods depends on reliability and closeness of the initial guess. This may create problems while dealing with the large scale scenes. The methods with level sets can work with arbitrary topologies and do not have restriction on the placement of cameras.

IMAGE SPACE APPROACH

Image space approach for 3D reconstruction is based on matching the features between image pairs robustly. These methods estimate the depth for each reference image entity (pixels, lines, windows, or segments) in the 3D space or along the corresponding epipolar lines and compute a set of depth maps. Further, a single consistent 3D scene interpretation is obtained by applying a consistency constraint to the set of depth maps and merging those into a 3D scene. The depth map merging approaches provide flexibility and superior performance (Seitz *et al*., 2006; Shen, 2012). The accuracy of 3D reconstruction is governed by the quality of depth maps. Traditionally, the depth maps are estimated in a discrete manner. Continuous variational depth map estimation overcomes the errors in discretization and reduces the memory consumption (Liu *et al*., 2009). High accuracy and robustness of the depth map merging algorithm are achieved by using high quality oriented 3D point clouds. Bundle optimization (Li *et al*., 2010) or expansion based depth estimation (Song *et al*., 2010) is used for this purpose. Delaunoy & Pollefeys (2014) proposed a bundle optimization of photo-metric re-projection error for collective optimization of the 3D shape and camera parameters. The method proposed by Campbell *et al*. (2008) forces a spatial consistency constraint on the discrete label optimization to recover the true depth from multiple depth maps.

This improves the performance of the reconstruction method. The approach based on binocular stereo proposed in the work of Bradley *et al.* (2008) uses an adaptive point based filtering, to filter the merged point clouds, and generates a surface mesh with good quality. A number of structured light depth cameras (SLDC) can be used for generation of good quality depth maps. An algorithm for merging such depth maps correlates multiple projectors and infrared images to recover the depth for non-overlapping and overlapping regions effectively (Wang *et al*., 2012). This helps to improve the quality of reconstruction. In case of the objects with sparse visible texture, time-of-flight (TOF) depth sensors are used to obtain the depth maps (Kim *et al*., 2009). Some methods extract geometric information in the pictures without using any assumption regarding the reflectance properties of the object and use this information in robust estimation of the depth maps (Zhou *et al*., 2013). RBF is used further to merge these depth maps (Lambert & Hebert, 2009).

Stability based fusion technique and confidence based fusion technique fail if true depth is not a part of the depth candidate set for a pixel. In large scale 3D reconstruction, pixels can be allowed to have multiple candidate depth values to overcome these limitations. Further, uncertainties due to mismatching and incorrect estimation of 3D coordinates are used to regulate the fusion process (Hu & Mordohai, 2012). In case of wide baseline image pairs, the estimation of dense depth map is done by using expectation minimization based algorithms (Tola *et al*., 2010). Typical image based fusion process regularizers such as total variation (TV) may lead to staircase artifacts. The regularizer for variational stereo based on the geometry of surface overcomes this drawback (Graber *et al*., 2015). A variational approach based on the prediction error minimization for elastic 3D motion estimation and multi-view stereo vision computes a matching score by matching the predicted images with the input images (Pons *et al*., 2007). As against the conventional methods, this approach overcomes the problems related to partial occlusion and projective distortion effectively. A variational approach for simultaneous estimation of the scene flow and structure uses the available multi-view information (Basha *et al*., 2013).

Reconstruction methods for the urban images demand automatic modeling of the scene. To solve this problem, these methods use shape priors at the initial stage and use this information further to reconstruct the scene (Labatut *et al*., 2009). The range data obtained from a range video gives direct information about the geometry of object surface (Zhou *et al*., 2013). However, these methods suffer from high frequency noise and quantization artifacts. These errors can be overcome by using volumetric registration approach for optimization. More accurate and reliable 3D reconstruction from higher resolution images is a challenging problem. The method proposed by Kim *et al*. (2013) reconstructs a complex scene with the details from densely sampled 3D light fields. Real time reconstruction of single 3D mesh and surface is possible by using Zippering algorithm along with iterative closest point (ICP) framework for real time merging of the point clouds (Alexiadis *et al*., 2013). Multi-view stereo (MVS) methods need photo-consistency computations for refinement of the 3D model. This results in a huge number of operations. The number of operations depends on the number of views and size of the model. To improve speed and performance, the method based on PatchMatch approach (Uh & Byun, 2016) uses random search and propagation for finding nearest-neighbor correspondences between the patches.

Image space approaches are more suitable for real time 3D rendering. This is because of less computation time, which depends on resolution of images and not on the scale/geometry of the scene/object. This also makes these methods more suitable for GPU implementations. Methods based on the depth map merging are more flexible and well suited for large scale reconstruction. The methods with dense stereo algorithms handle occlusions as against the methods based on feature detection and matching. These methods do not need background removal and segmentation of the object in the views. This makes the methods more robust against background noise. Image space methods rely on solving the correspondence problem. Hence, the success of this step affects the accuracy of these methods. In presence of the noise, triangulation step in these methods reduces the accuracy of the reconstruction.

EXTRACTING FEATURE POINTS AND FITTING A SURFACE

This class of approaches consists of the methods that extract a set of feature points and fit the surface to the extracted points (Seitz *et al*., 2006). Many methods use scale invariant feature transform (SIFT) to detect the features in the images (Divyalakshmi & Vaithiyanathan, 2016). The surface reconstruction methods use the initial surface fitted to the data, in order to reflect the surface represented in the point cloud. In case of unorganized point cloud, the prior information about the shape of the object is required for fitting the surface dynamically to the unorganized point cloud (Nurzynska, 2009). The point cloud can also be obtained by extracting a set of points representing the occluding and texture edges (Liu *et al*., 2008). Photo-consistency of 3D points based on squared Euclidean RGB distance between a pair of image points can also be used for extracting the point cloud (Salvador & Casas, 2010). Feature based methods face challenges while reconstructing the featureless objects. With the additional cues and prior information, these challenges can be overcome (Ley *et al*., 2016; Bao *et al*., 2013). The methods using RGB-D cameras obtain 3D point cloud using depth maps and construct a surface using truncated sign distance fields (TSDFs) (Henry *et al*., 2016). 3D reconstruction methods using low cost RGB-D cameras such as Microsoft Kinect face challenges in handling repeated texture regions. The methods use visual and geometrical features along with the structure from motion technique to recover the missing geometry (Wang *et al*., 2014). A real time implementation of a similar approach based on RGB-D cameras is presented in the work of Zollhöfer *et al*. (2014). Online reconstruction methods based on TSDF define the error function using the error from corresponding sparse feature points. They obtain correspondences between the currently acquired images and the earlier images to create an error function independent of the model (Bylow *et al*., 2016).

Direct surface reconstruction is difficult from the sparse 3D feature points since they are missing in large areas and also irregularly distributed. To overcome this problem, the methods reconstruct small surface patches. Further, the patches, which are consistent with their neighborhood, are used to obtain the entire surface of the object (Zeng *et al*., 2004). In such patch based methods, the polygon mesh is used to represent these patches (Furukawa & Ponce, 2010). Huang *et al*. (2016) proposed the sparse patch based method using a monocular camera. The method considers the depth obtained from the mapping thread to control the tracking process. Instead of using such sparse 3D points, the free space volume obtained from stereo can also be used for generating the surfaces (Taylor, 2003). The method uses 3D results obtained from stereo to determine the structure of input scene rather than applying constraints based on viewpoint consistency. MVS methods focus on reconstructing all the details of the object under reconstruction. When such details are not required, the computation time can be reduced by using an approach proposed by Bodis *et al*. (2015). The method optimizes the dense depth maps over sparse ground control points (GCP) to remove the need of stereo depth estimation. This results in reduction of the computation time.

The reconstructed scene if modeled as probability distributions in 3D space allows use of features such as the points on edges, which cannot be matched one to one or whose locations cannot be determined precisely (Teney & Piater, 2012). Image gradient based techniques are used to extract edge orientations needed for obtaining such features. Feature based MVS methods can be extended to dynamic scenes such as 3D videos. However, solving the stereo correspondence becomes a difficult task in such cases. Limitations such as incompleteness of the reconstructed model and weak photo-consistency of stereo reduce the accuracy of reconstruction. These limitations can be overcome by fusing a multi-view structure from motion with robust 3D features obtained from MVS (Tung *et al*., 2009). Methods that need the knowledge of a 3D scene and its approximate geometry cannot be used for a large scale scene reconstruction. The method based on matching of key interest points for creating a quasi-dense 3D point cloud and labeling of Delaunay tetrahedral as 'occupied' or 'empty' overcomes these drawbacks (Labatut *et al*., 2007).

The methods based on feature detection and matching are computationally more efficient. They need less memory as compared to voxel based methods. However, these methods face challenges in reconstruction of the featureless objects such as objects with shiny, textureless, and smooth surface. View space conversion leads to the introduction of visual artifacts during the image rendering process.

COMPARISON OF RECONSTRUCTION METHODS

This section presents a comparison of different reconstruction methods on the basis of the performance and key parameters. Table 1 shows the performance evaluation of different methods on DinoRing dataset with 48 views and TempleRing dataset with 47 views (Seitz *et al*., 2006) provided by Middlebury using two performance parameters, completeness and accuracy.

Table 1. Performance evaluation of the methods using Middlebury dataset.

Method	DinoRing (48 views)		TempleRing (47 views)	
	Completeness %	Accuracy mm	Completeness %	Accuracy mm
Furukawa (Furukawa & Ponce, 2010)	99.3	0.33	99.1	0.57
Kolev (Kolev & Cremers, 2008)	99.4	0.43	97.8	0.72
Kostrikov (Kostrikov *et al*., 2014)	99.6	0.35	99.1	0.57
Bradley (Bradley *et al*., 2008)	97.6	0.39	98.1	0.57
Campbell (Campbell *et al*., 2011)	-	-	99.4	0.48
Vu (Vu *et al*., 2009)	-	-	99.8	0.45
Zaharescu (Zaharescu *et al*., 2011)	98.6	0.42	99.2	0.55
Uh (Uh & Byun, 2016)	97.3	0.32	96.4	0.51
Hernandez (Hernandez *et al*., 2008)	97.9	0.45	99.5	0.52
Vogiatzis (Vogiatzis *et al*., 2007)	96.7	0.49	96.2	0.76
Pons (Pons *et al*., 2007)	99.0	0.55	99.5	0.60
Kolmogorov (Boykov & Kolmogorov, 2004)	85.7	2.80	90.4	1.86

Completeness of the reconstruction is determined by computing a percentage of the reconstructed model with respect to the ground truth model. Accuracy of the reconstruction is obtained by computing the distance between reconstructed model points and nearest points on the ground truth model. It is expressed in millimeters. Lower values of this parameter indicate high accuracy. It is observed that most of the methods considered here are able to reconstruct the complete object model. Most of the methods achieve completeness more than 95%. The method based on probabilistic labeling cost (Kostrikov *et al*., 2014) produces the most complete model as compared to the other methods. The exception is Kolmogorov's method. The construction consists of the holes. Hence, it exhibits a lower completeness value. It is observed that the accuracy of many reconstruction methods is up to the mark. These methods exhibit accuracy in sub-millimeter. Uh's method (Uh & Byun, 2016) outperforms other methods with almost 90% points within 0.32mm of the ground truth model for DinoRing dataset, whereas Vu's method (Vu *et al*., 2009) exhibits the highest accuracy with 90% points within 0.45mm of the ground truth model for TempleRing dataset. Methods by Furukawa, Pons, and Kostrikov show the best performance for both datasets. Tables 2 and 3 provide the information about the widely used toolboxes/open source software and datasets in the field of multi-view 3D reconstruction, respectively. Methods presented in different approaches are also compared on the basis of their important key features. Tables 4 and 5 present the comparison between the methods discussed under the subsection entitled "surface extraction based on cost function computed on volume". Tables 6 and 7 compare the methods based on the surface evolution approach. Tables 8, 9, and 10 compare the methods based on the image space approach, whereas tables 11 and 12 compare the methods based on the feature point extraction for surface fitting. In tables 4 to 12, "√" (tick) indicates that the corresponding parameter is satisfied by the method described in the paper, whereas "X" (cross) indicates that the corresponding parameter is not satisfied by the method described in the paper. "NA" indicates that the corresponding parameter is not addressed in the paper.

Table 2. Tool boxes widely used in the field of multi-view 3D reconstruction.

Tool box/ open source s/w	Brief description
VisualSFM (Wu, 2013)	GUI application for 3D reconstruction using structure from motion (SFM)
MVE (Fuhrmann *et al*., 2014)	Implementations for structure from motion, multi-view stereo, and surface reconstruction
OpenMVG (Pierre Moulon *et al*., 2013)	Library for multiple view geometry
SFMToolbox (Rabaud, 2009)	Structure from motion tool box developed in MATLAB
Bundler (Snavely *et al*., 2007)	Open source SFM system for sparse point clouds
PMVS2 and CMVS (Furukawa & Ponce, 2010)	Multi-view stereo framework with patch based (PMVS) and clustering (CMVS)
CMPMVS (Jancosek & Pajdla, 2010)	MVS framework
Matlab Functions for Multiple View Geometry (Hartley *et al*., 2004)	Matlab functions developed for multiple view geometry by Andrew Zisserman
Camera Calibration Tool Box (Bouguet & Perona, 2009)	Camera calibration tool box developed by J-Y Bouguet and available on Caltech University website

Table 3. Datasets widely used in the field of multi-view 3D reconstruction.

Datasets	**Brief description**
Ford Car Dataset (Pandey *et al*., 2011)	Images with Lidar points
StrechaDenseMVS (Strecha *et al*., 2008)	Dataset for dense MVS include fountain-p11, castle and Herz-Jesu-p8 datasets
Leuven Stereo Scene (Leibe *et al*., 2007)	Stereo dataset with scene and depth
Car and bottle (Savinov *et al*., 2015)	Dataset of images with 3D object shape priors
Middlebury Stereo datasets–2005 (Scharstein & Pal, 2007)	Stereo dataset
Middlebury Stereo datasets–2006 (Scharstein & Pal, 2007)	Stereo dataset with conditional random field(CRF) framework
Middlebury Stereo datasets–2014 (Scharstein *et al*., 2014)	High resolution stereo dataset
Temple and Dino datasets (Seitz *et al*., 2006)	Multi-view stereo dataset with ground truth
Kermit, Citywall, Der Hass datasets (Fuhrmann *et al*., 2014)	Images with VisualSFM bundle file
3D scene dataset (Li *et al*., 2015)	Dataset with sequence having RGB-D frames

Table 4. Comparison of methods extracting surface based on cost function computed on volume.

Paper Parameter	**Adipran-ata & Soo, 2007**	**Slem-bro-uck et al., 2014**	**Brisc, 2004**	**Zao & Xiao, 2005**	**Feldm-ann et al., 2010**	**Boykov & Kolmogo-rov, 2004**	**Vogiatzis et al., 2007**	**Perez et al., 2012**
Calibrated images	√	√	X	√	√	√	√	√
Real objects/scene	√	√	√	√	√	√	√	√
Optimization	X	√	NA	NA	NA	√	√	√
Mean chart based colour consistency	√	X	X	X	X	X	X	X
HLS Space	X	X	X	√	X	X	X	X
Use epipolar geometry	X	√	X	X	X	√	X	√
Auto-calculation of bounding box	X	X	X	X	√	√	X	X
Correlation based cost function	X	X	X	X	X	√	X	√
Self-learning for voxel occlusions	X	√	X	X	X	X	X	X
GPU implementa-tion	X	X	X	X	X	X	X	√

Table 5. Comparison of methods extracting surface based on cost function computed on volume.

Paper Parameter	Franco & Boyer, 2009	Hernan-dez et al., 2008	Bottino & Laurentini, 2000	Shin & Tjahjadil, 2008	Monten-egro et al., 2006	Tran et al., 2006	Camp-bell et al., 2011	Ulu-soy et al.,2016
Polygon mesh	√	√	√	X	X	X	X	X
Marching cubes	X	X	X	√	X	X	X	X
Signed OCTrees	X	X	X	X	√	X	X	√
Less computation complexity	√	NA	NA	√	NA	√	√	√
Optimization	x	√	NA	NA	X	√	X	√
Good accuracy	NA	√	NA	√	NA	√	√	√
Estimation based on epipolar geometry	√	X	X	X	X	X	√	X
Textureless shiny object	X	√	X	X	X	X	√	√
Unknown relative viewpoints	X	X	√	X	X	X	X	X
Handles imperfect silhouettes	X	X	X	√	X	√	√	NA
Graph cut	X	X	X	X	X	√	√	X
Real data	√	√	√	√	√	√	√	√
MRF framework	X	X	X	X	X	X	√	√
Voxels	X	X	X	X	X	√	√	√

Table 6. Comparison of methods based on surface evolution approach.

Paper Parameter	Goldluecke & Magnor, 2004	Kolev & Cremers, 2008	Zaharescu et al., 2011	Vu et al., 2009	Khan et al., 2007	Ilic & Fua, 2006	Blaha et al., 2016	Nehab et al., 2008	Park et al., 2013
Calibrated images	√	√	X	√	√	X	√	√	X
Real dataset	√	√	√	√	√	√	√	√	√
Computation complexity	√	√	√	NA	NA	√	√	√	√
Optimization	√	√	√	√	√	√	√	√	√
Good accuracy	√	√	√	√	X	X	√	X	√
convex function	√	√	X	X	X	X	√	X	X
Dynamic scene	√	X	√	X	X	X	X	X	X

Temporal coherence for photo consistency	√	X	X	X	X	X	X	X	X
Volumetric graph cuts	X	√	X	X	X	X	X	X	X
Silhouette-stereo fusion	X	√	√	X	X	X	X	X	X
convex constraint	X	√	√	X	X	X	√	X	X
TansforMesh	X	X	√	X	X	X	X	X	X
Large scale data	X	X	X	√	X	X	√	X	X
Topology adaptive	X	X	√	√	X	X	X	X	X
Weighted RBF	X	X	X	X	X	X	X	X	X
Mesh parameterization	X	X	X	X	X	X	√	√	√

Table 7. Comparison of methods based on surface evolution approach.

Paper Parameter	**Ju et al., 2009**	**Cremers & Kolev, 2011**	**Lhuillier & Quan, 2005**	**Nghiem et al., 2010**	**Chang et al., 2007**	**Dainese et al., 2005**	**Savinov et al., 2015**	**Kostrikov et al., 2014**
Calibrated images	√	X	X	√	√	√	√	√
Real dataset	√	√	√	√	X	√	√	√
Computation complexity (Less)	√	√	NA	√	NA	√	√	NA
Optimization	√	√	√	√	√	√	√	√
Good accuracy	NA	√	√	√	√	NA	√	√
Deformable model	√	X	X	X	X	√	X	X
Convex function	X	√	X	X	X	X	X	X
Function based on 2D & 3D data points	X	X	√	X	X	X	X	√
RBF	X	X	X	√	X	X	X	X
Least squared error	X	X	X	X	√	X	X	X
Wide base line	X	√	X	√	X	X	X	√
Labeling cost/ smoothing term	X	√	X	X	X	X	√	X

Table 8. Comparison of methods based on image space approach.

Paper Parameter	**Campbell et al., 2008**	**Alexiadis et al., 2013**	**Pons et al., 2007**	**Wang et al., 2012**	**Labatut et al., 2009**	**Bradley et al., 2008**	**Hu & Mordoh-ai, 2012**	**Zhou et al., 2013**
Calibrated images	√	X	X	√	√	√	√	√
Real dataset	√	√	√	√	√	√	√	√
Computation complexity (less)	√	√	NA	NA	NA	√	NA	NA
Optimization	√	√	√	X	√	√	X	√
Wide base line	√	√	√	√	X	√	√	X
Dynamic scene	X	√	√	X	X	X	X	X
Spatial consistency constraint	√	X	X	X	X	X	X	X
RGB depth maps	X	√	X	X	X	X	X	X
Zippering algorithm	X	√	X	X	X	X	X	X
Prediction error	X	X	√	X	X	X	X	X
Structured light based cameras	X	X	X	X	√	X	X	X
Adaptive point based filtering	X	X	X	X	X	√	X	X
Multiple candidate depths per pixel	X	X	X	X	X	X	√	X
Range data	X	X	X	X	X	X	X	√

Table 9. Comparison of methods based on image space approach.

Paper Parameter	**Graber et al., 2015**	**Basha et al., 2013**	**Kim et al., 2013**	**Uh & Byun, 2016**	**Shen, 2012**
Calibrated images	√	√	√	√	X
Real dataset	√	√	√	√	√
Computation complexity (less)	√	X	√	√	√
Optimization	√	√	X	√	X
Wide baseline	X	X	√	√	√
Dynamic scene	X	√	X	X	X
Spatial consistency constraint	X	X	X	√	X
RGB depth maps	X	√	√	X	X
Adaptive point based filtering	X	X	√	√	X
Multiple candidate depths/pixel	X	X	X	√	X
Use of range data	√	X	X	X	X
Variational Approach	√	√	X	X	X

Table 10. Comparison of methods based on image space approach.

Paper Parameter	Kim et al., 2009	Tola et al., 2010	Liu et al., 2009	Song et al., 2010	Li et al., 2010	Lambert & Hebert, 2009	Zhou et al., 2013	Delaunoy et al., 2014
Calibrated images	√	√	X	√	X	√	√	√
Real dataset	√	√	√	√	√	√	√	√
Computation complexity (less)	NA	√	NA	X	√	X	X	X
Optimization	√	√	√	√	√	√	√	√
Sensor fusion for finding depth	√	X	X	X	X	X	X	X
DAISY descriptor	X	√	X	X	√	X	X	X
Continuous depth map estimation	X	X	√	X	X	X	√	√
Expansion based approach	X	X	X	√	x	X	X	X
Bundle optimization for robust depth map merging	X	X	X	X	√	√	X	√
Depth map based on silhouette and epipolar line	X	X	√	X	X	√	X	X
Wide base line	X	√	√	X	√	X	X	X
Dynamic scene	√	X	X	X	X	X	X	X
Use of reflectance properties	X	X	X	X	X	√	√	X

Table 11. Comparison of methods that extract a set of feature points and then fits the surface on it.

Paper Parameter	Labatut et al., 2007	Taylor, 2003	Zeng et al., 2004	Teney & Piater, 2012	Tung et al., 2009	Furukawa & Ponce, 2010	Liu et al., 2008	Nurzynska, 2009	Salvador & Casas, 2010
Calibrated images	√	X	X	X	X	X	√	NA	√
Real dataset	√	√	√	√	√	√	√	√	√
Computation complexity	√	NA	NA	X	NA	X	X	X	X
Optimization	√	X	√	X	√	√	√	X	X
Initialization (visual hull)	X	X	X	X	√	X	√	X	√
Wide base line	√	X	X	X	√	√	X	√	√
Feature points (texture & occluding edges)	X	X	X	√	X	x	√	X	√
Good accuracy	√	NA	NA	√	√	√	√	X	√

Unorganized point cloud	X	X	X	X	√	X	X	√	X
Large scale scenes	√	X	X	X	√	X	X	X	X
Feature point matching	√	√	√	√	√	√	√	√	√
free space volume based features	X	√	X	X	X	X	X	X	X
Fusion of narrow & wide baseline	X	X	X	X	√	X	X	X	X
Probabilistic approach	X	X	X	√	√	X	X	X	X

Table 12. Comparison of methods that extract a set of feature points and then fits surface on it.

Paper Parameter	**Wang et al., 2014**	**Huang et al., 2016**	**Henry et al., 2016**	**Bylow et al., 2016**	**Bao et al., 2013**	**Bodis et al., 2015**
calibrated images	√	√	√	√	X	√
Real dataset	√	√	√	X	√	√
Computation complexity	√	√	X	X	NA	√
Optimization	√	√	√	X	X	√
Initialization (visual hull)	X	X	X	X	X	X
Wide base line	X	X	√	√	√	√
Feature points (occluding & texture edges)	X	X	X	X	X	√
Good accuracy	√	√	√	√	X	NA
Dynamic surface (unorganized point cloud)	X	X	X	X	NA	X
Large scale scenes	X	X	√	√	√	√
Feature point matching	√	√	√	√	√	√
features (free space volume)	X	X	X	X	X	X
Fusion of narrow & wide baseline	X	X	X	√	X	X
Probabilistic approach	X	X	X	X	X	X
RGB depth cameras	√	X	√	X	X	√
Truncated Signed distance function	√	X	√	X	X	X

DISCUSSION AND CONCLUSION

This paper presents a review and comparison of the latest 3D reconstruction methods developed by computer vision researchers. Although it is very difficult to classify these methods, they are grouped on the basis of their common features. Different reconstruction methods are compared on the basis of their performance and key parameters. The results of the efficient and accurate volumetric 3D reconstruction method developed by the authors are also discussed. This will help researchers to understand the state of the art in this field.

Solving a correspondence problem is one of the complex tasks in itself. On the other hand, multi-view reconstruction methods that operate in 3D space such as voxel based methods do not need to solve the complex correspondence problem. The volumetric methods may fail to reconstruct non-Lambertian objects. The beauty of voxel based methods lies in the fact that they provide physical parameters close to those of the actual scene. Methods based on the image space approach use estimation techniques to provide better accuracy than the volumetric methods. Applications in the medical field demand more accuracy. Hence, the approaches can be combined to take benefits of both.

Development of multi-view 3D reconstruction methods has experienced a tremendous growth in recent years. Still, these methods need improvement in many aspects such as reduction in computation complexity, implementation simplicity, suitability for large scale and dynamic scenes, suitability for real time applications, and so forth. The literature shows that many reconstruction methods are basically designed for small baseline and indoor applications. Emphasis can be given on making these methods suitable for outdoor applications. Most of the methods model static scene only. The methods can be modified to make them suitable for dynamic scene reconstruction. Although 3D reconstruction methods using MVS can reconstruct good quality large scale scenes/models, they do not produce good quality textured results. Efforts can be made to improve these methods to obtain good quality of texture. More emphasis should be given on making these complex algorithms realizable on the hardware platforms such as GPU. In volumetric reconstruction methods, the accuracy of reconstruction depends upon calibration of cameras. The calibration process can be automated to make the method simple and improve the overall performance. Reconstruction methods mainly face challenges while dealing with the complex scenes and objects with textureless surface. With additional cues/priors, the methods can be modified to deal with such objects. Transparent objects like fire or objects with reflecting surface like mirrors are difficult to reconstruct. More efforts are needed in this direction. There is a scope to improve the accuracy of image space methods by improving the accuracy of depth maps. The scope of 3D reconstruction is increasing day by day. It is challenging to make these methods more suitable for satellite imagery and geographic information system (GIS) applications where high resolution images are to be dealt with.

Acknowledgment

The authors wish to thank All India Council for Technical Education (AICTE), New Delhi, for funding the proposed work under Research Promotion Scheme (RPS) [Sanction no. 8023/RID/RPS-160 (PVT)/2011-12].

References

Adipranata, R. & Soo, Y. 2007. 3D object reconstruction using voxel colouring with mean charts colour consistency checking. Proceedings of International Conference in Advances in Computer Science and Technology. Phuket, Thailand.

Alexiadis, D.S., Zarpalas, D. & Daras P. 2013. Real-time, full 3-D reconstruction of moving foreground objects from multiple consumer depth cameras. IEEE Transactions on Multimedia, **15**(2): 339-358.

Bao, S., Chandraker, Y.M., Lin, Y. & Savarese, S. 2013. Dense object reconstruction with semantic priors. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, USA.

Basha, T., Moses, Y. & Kiryati, N. 2013. Multi-view scene flow estimation: A view centered variational approach. International Journal of Computer Vision, **101**(1): 6–21.

Blaha, M., Vogel, C., Richard, A., Wegner, J., Pock, T. & Schindler, K. 2016. Large-scale semantic 3D reconstruction: An adaptive multi-resolution model for multi-class volumetric labeling. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). Las Vegas, Nevada, USA.

Bódis-Szomorú, A., Riemenschneider, H. & Gool L.V. 2015. Superpixel meshes for fast edge-preserving surface reconstruction. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence, Rhode Island.

Bottino, A. & Laurentini A. 2003. Introducing a new problem: Shape-from-silhouette when the relative positions of the viewpoints is unknown. IEEE Transactions on Pattern Analysis and Machine Intelligence, **25**(11): 1484-1493.

Bouguet, J. & Perona, P. 1999. 3D Photography using shadows in dual-space geometry. International Journal of Computer Vision (IJCV), **35**(2): 129-149

Boykov, Y. & Kolmogorov V. 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, **26**(9): 1124-1137.

Bradley, D., Boubekeur T. & Heidrich, W. 2008. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, Alaska, USA.

Brisc F. 2004. Multi-resolution volumetric reconstruction using labeled regions. Proceedings of 6th IEEE Southwest Symposium on Image Analysis and Interpretation. Nevada, USA.

Bylow, E., Olsson, C. & Kahl, F. 2016. Robust online 3D reconstruction combining a depth sensor and sparse feature points. Proceedings of International Conference on Pattern Recognition. Cancun, Mexico.

Campbell, N.D.F., Vogiatzis, G., Hernandez, C. & Cipolla R. 2008. Using multiple hypotheses to improve depth-maps for multi-view stereo. Proceedings of 10th European Conference on Computer Vision: Part I (ECCV'08). Marseille, France.

Campbell, N.D.F., Vogiatzis, G., Hernandez, C. & Cipolla, R. 2011. Automatic object segmentation from calibrated images. Proceedings of European Conference on Visual Media Production (CVMP). Washington, DC, USA.

Chang, J.Y., Lee, K.M. & Lee, S.U. 2007. Multi-view normal field integration using level set methods. Proceedings IEEE International Conference on Computer Vision and Pattern Recognition. Minneapolis, USA.

Cremers, D. & Kolev K. 2011. Multi-view stereo and silhouette consistency via convex functionals over convex domains. IEEE Transactions on Pattern Analysis and Machine Intelligence, **33**(6): 1161-1174.

Dainese, G., Marcon, M., Sarti, A. & Tubaro, S. 2005. 3D object modeling with a voxel-set carving approach. Proceedings of IEEE International Conference on Image Processing (ICIP 05). Geneva, Italy.

Delaunoy, A. & Pollefeys, M. 2014. Photometric bundle adjustment for dense multi-view 3D modeling. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Columbus, Ohio, USA.

Dinh, H.Q., Turk, G. & Slabaugh, G. 2002. Reconstructing surfaces by volumetric regularization using radial basis functions. IEEE Transactions on Pattern Analysis and Machine Intelligence, **24**(10): 1358-1371.

Divyalakshmi K. & Vaithiyanathan V. 2016. Image registration techniques based on the scale invariant feature transform. IETE Technical Review, 1-8.

Dyer C. 2001. Volumetric scene reconstruction from multiple views. Foundations of Image Understanding. Pp. 469-489, Springer, USA.

Feldmann, T., Brand, K. & Worner, A. 2010. Enhancing voxel carving by capture volume calculations. Proceedings of 17th IEEE International Conference on Image Processing (ICIP), Hong Kong.

Franco, J. & Boyer E. 2009. Efficient polyhedral modeling from silhouettes. IEEE Transactions on Pattern Analysis and Machine Intelligence, **31**(3): 414-427.

Fuhrmann, S., Langguth, F. & Goesele, M. 2014. MVE: A multi-view reconstruction environment. Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage (GCH '14). Geneva, Italy.

Furukawa, Y. & Ponce, J. 2010. Accurate, dense, and robust multi-view stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence, **32**(8): 1362-1376.

Goldluecke, B. & Magnor, M. 2004. Space-time isosurface evolution for temporally coherent 3D reconstruction. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2004). Washington, DC, USA.

Graber, G., Balzer, J., Soatto, S. & Pock T. 2015. Efficient minimal-surface regularization of perceptive depth maps in variational stereo. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA.

Henry, P., Krainin, M., Herbst, E., Ren, X. & Fox, D. 2016. Dense 3D reconstruction combining depth and RGB information. Journal of Neurocomputing, **175**: 644-651

Hernandez, C., Vogiatzis, G. & Cipolla, R. 2008. Multi-view photometric stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, **30**(3): 548-554.

Hu, X. & Mordohai, P. 2012. Least commitment, viewpoint-based, multi-view stereo. Proceedings of Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT' 12). Tokyo, Japan.

Huang, X., Fan, L., Zhang, J., Wu, Q. & Yuan, C. 2016. Real time complete dense depth reconstruction for a monocular camera. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA.

Ilic, S. & Fua, P. 2006. Implicit meshes for surface reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, **28**(2): 328-333.

Jancosek, M. & Pajdla, T. 2011. Multi-view reconstruction preserving weakly-supported surfaces. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, Rhode Island.

Hartley, R. & Zisserman, A. 2003. Multiple view geometry in computer vision, 2nd ed. Cambridge University Press, NY.

Ju, X., Siebert, J.P., Khambay, B.S. & Ayoub, A.F. 2009. Self-correction of 3D reconstruction from multi-view stereo images. Proceedings of IEEE 12th International Conference on Computer Vision Workshops (ICCV workshops). Kyoto, Japan.

Khan, S.M., Yan, P. & Shah, M. 2007. A homographic framework for the fusion of multi-view silhouettes. Proceedings of IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil.

Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A. & Gross, M. 2013. Scene reconstruction from high spatio-angular resolution light fields. ACM Transactions on Graphics, **32** (4): 73:1-73:12.

Kim, Y.M., Theobalt, C., Diebel, J., Kosecka, J., Miscusik, B. & Thrun, S. 2009. Multi-view image and ToF sensor fusion for dense 3D reconstruction. Proceedings of IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops). Kyoto, Japan.

Kimura, M., Saito, H. & Kanade, T. 1999. 3D voxel construction based on epipolar geometry. Proceedings of IEEE International Conference on Image Processing (ICIP-99). Kobe, Japan.

Kolev, K. & Cremers, D. 2008. Integration of multi-view stereo and silhouettes via convex functionals on convex domains. Proceedings of European Conference on Computer Vision (ECCV). Marseille, France.

Kostrikov, I., Horbert, E. & Leibe, B. 2014. Probabilistic labeling cost for high-accuracy multi-view reconstruction. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Ohio, USA.

Labatut, P., Pons, J.P. & Keriven, R. 2007. Efficient multi-view reconstruction of large-scale scenes using interest points, Delaunay triangulation and graph cuts. Proceedings of IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil.

Labatut, P., Pons, J.P. & Keriven, R. 2009. Hierarchical shape-based surface re-construction for dense multi-view stereo. Proceedings of IEEE International Conference on Computer Vision Workshops (ICCV), Kyoto, Japan.

Lambert P. & Hebert P. 2009. Robust multi-view stereo without matching. Proceedings of IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops). Kyoto, Japan.

Leibe, B., Cornelis, N., Cornelis, K. & Van Gool, L. 2007. Dynamic 3D scene analysis from a moving vehicle. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, USA.

Ley, A., Hänsch, R. & Hellwich, O. 2016. Reconstructing white walls: Multi-view, multi-shot 3D reconstruction of textureless surfaces. Proceedings of ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Prague, Czech Republic.

Lhuillier, M. & Quan, L. 2005. A quasi-dense approach to surface reconstruction from uncalibrated images. IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**(3): 418-433.

Li, J., Li, E., Chen, Y., Xu, L. & Zhang, Y. 2010. Bundled depth-map merging for multi-view stereo. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA.

Liu, S., Kang, K., Tarel, J.P., & Cooper, D.B. 2008. Free-form object reconstruction from silhouettes, occluding edges and texture edges: A unified and robust operator based on duality. IEEE Transactions on Pattern Analysis and Machine Intelligence, **30**(1): 13-146.

Liu, Y., Cao, X., Dai, Q. & Xu, W. 2009. Continuous depth estimation for multi-view stereo. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida.

Matusik, W., Buehler, C., Raskar, R., Gortler, S. & McMillan, L. 2000. Image based visual hulls. Proceedings of 27th Annual Conference on Computer Graphics and Interactive Techniques. New Orleans, Louisiana, USA.

Montenegro, A.A., Velho, L., Carvalho, P. & Sossai, J.J. 2006. Polygonization of volumetric reconstructions from silhouettes. Proceedings of 19th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'06). Manaus, Amazonas, Brazil.

Moretto, N., Cugola, N. & Frezza R. 2003. Volumetric reconstruction by random sampling. Proceedings of 42nd IEEE Conference on Decision and Control. Hawaii, USA.

Moulon, P., Monasse, P. & Marlet, R. 2013. Global fusion of relative motions for robust, accurate and scalable structure from motion. Proceedings of IEEE International Conference on Computer Vision. Sydney, Australia.

Nehab, D., Weyrich, T. & Rusinkiewicz, S. 2008. Dense 3D reconstruction from specularity consistency. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska, USA.

Nghiem, V.N., Cai, J. & Zheng, J. 2010. Rate-distortion optimized progressive 3D reconstruction from multi-view Images. Proceedings of 18th Pacific Conference on Computer Graphics and Applications (PG). Hangzhou, China.

Nurzynska, K. 2009. Constrained dynamical surface reconstruction from unorganized point cloud. Proceedings of IEEE EUROCON'09. St. Petersburg, Russia.

Pandey, G., McBride, J. & Eustice, R. 2011. Ford campus vision and Lidar data set. International Journal of Robotics Research, **30**(13): 1543-1552.

Park, J., Sinha, S.N., Matsushita, Y., Tai, Y.W. & Kweon, I.S. 2013. Multi-view photometric stereo using planar mesh parameterization. Proceedings of IEEE International Conference on Computer Vision. Sydney, Australia.

Pons, J.P., Keriven, R. & Faugeras, O. 2007. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. International Journal of Computer Vision, **72**(2): 179-193.

Perez, J., Aledo, P. & Sanchez, P. 2012. Real-time voxel-based visual hull reconstruction. Microprocessors and Microsystems, **36**(5): 439–447.

Rabaud, V. & Belongie, S. 2009. Linear embeddings in non-rigid structure from motion. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami, Florida, USA.

Salvador, J. & Casas, J.R. 2010. Photo consistent surfaces from a sparse set of viewpoints. Proceedings of IEEE 17th International Conference on Image Processing. Hong Kong.

Savinov, N., Ladický, L., Häne, C. & Pollefeys, M. 2015. Discrete optimization of ray potentials for semantic 3 D reconstruction. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA.

Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nesic, N., Wang, X. & Westling, P. 2014. High-resolution stereo datasets with subpixel-accurate ground truth. Proceedings of German Conference on Pattern Recognition (GCPR 2014). Munster, Ireland.

Seitz, S.M., Curless, B., Diebel, J., Scharstein, D. & Szeliski, R. 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06). New York, USA.

Shen, S. 2012. Depth-map merging for multi-view stereo with high resolution images. Proceedings of 21st International Conference on Pattern Recognition (ICPR 2012). Tsukuba Science City, Japan.

Shin, D. & Tjahjadi, T. 2008. Local hull-based surface construction of volumetric data from silhouettes. IEEE Transactions on Image Processing, **17**(8): 1251-1260.

Slabaugh, G., Culbertson, B., Malzbender, T. & Schafer, R.2001. A survey of methods for volumetric scene reconstruction from photographs. Proceedings of the Joint IEEE TCVG and Eurographics Workshop, Volume Graphics 2001, Springer Computer Science, Stony Brook, New York, USA.

Slembrouck, M., Cauwelaert, D., Hamme, D., Haerenborgh, D., Hese, P., Veelaert, P. & Philips, W. 2014. Self-learning voxel-based multi-camera occlusion maps for 3D reconstruction. Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP). Lisbon, Portugal.

Snavely, N., Seitz, S. & Szeliski, S. 2008. Modeling the world from internet photo collections. International Journal of Computer Vision, **80**(2): 189-210.

Song, P., Wu, X., Wang, M.Y. & Wu, J. 2010. Expansion-based depth map estimation for multi-view stereo. Proceedings of International Conference on Intelligent Robots and System. Taipei International Convention Center, Taipei, Taiwan.

Strecha, C., VonHansen, G., Van Gool, L., Fua, P. & Thoennessen, U. 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Anchorage, Alaska, USA.

Taylor, C.J. 2003. Surface reconstruction from feature based stereo. Proceedings of IEEE International Conference on Computer Vision. Nice, France.

Teney, D. & Piater, J. 2012. Sampling-based multi-view reconstruction without correspondences for 3D edges. Proceedings of IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DimPVT), Oct.12. Zurich, Switzerland.

Tola, E., Lepetit, V. & Fua, P. 2010. DAISY: An efficient dense descriptor allied to wide-baseline stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, **32** (5): 815-830.

Tran S. & Davis L. 2006. 3D surface reconstruction using graph cuts with surface constraints. Proceedings of European Conference on Computer Vision: Part II (ECCV), 219-231.

Tung, T., Nobuhara, S. & Matsuyama T. 2009. Complete multi-view reconstruction of dynamic scenes from probabilistic fusion of narrow and wide baseline stereo. Proceedings of IEEE 12th International Conference on Computer Vision. Kyoto, Japan.

Uh Y. & Byun, H. 2016. Multi-view 3D reconstruction by random-search and propagation with view-dependent patch maps. Multimedia Tools and Applications, **75**: 1-18.

Ulusoy, A.O., Black, M.J. & Geiger, A. 2016. Patches, planes and probabilities: A non-local prior for volumetric 3D reconstruction. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, Nevada, USA.

Vogiatzis, G., Hernandez, C., Torr, P. & Cipolla, R. 2007. Multi-view stereo via volumetric graph-cuts and occlusion robust photo-consistency. IEEE Transactions on Pattern Analysis and Machine Intelligence, **29**(12): 2241-2246.

Vu, H., Keriven, R., Labatut, P. & Pons, J.P. 2009. Towards high-resolution large scale multi-view stereo. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009). Miami, Florida, USA.

Wang, J., Zhang, C., Zhu, W., Zhang, Z., Xiong, Z. & Chou, P.A. 2012. 3D scene reconstruction by multiple structured-light based commodity depth cameras. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Kyoto, Japan.

Wang, K., Zhang, G. & Bao, H. 2014. Robust 3D reconstruction with an RGB-D camera. IEEE Transactions on Image Processing, **23**(11): 4893 – 4906.

Wu, C. 2013. Towards linear-time incremental structure from motion. Proceedings of International Conference on 3D Vision (3DV 2013). University of Washington, Seattle, USA.

Wu, J., Deng, L. & Paul, A. 2015. 3D Terrain real time rendering method based on CUDA-OpenGL interoperability. IETE Technical Review, **32**(6): 471-478.

Zaharescu, A., Boyer, E. & Horaud, R. 2011. Topology-adaptive mesh deformation for surface evolution, morphing and multi-view reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, **33**(4): 823-837.

Zao, C. & Xiao W. 2005. A parallelizing and improving on voxel colouring technology for 3D reconstruction based on images. Proceedings of IEEE International Conference on Mechatronics and Automation. Canada.

Zeng, G., Paris, S., Quan, L., & Lhuillier, M. 2004. Surface reconstruction by propagating 3D stereo data in multiple 2D images. Proceedings of European Conference on Computer Vision. Prague.

Zhou, Z., Wu, Z. & Tan, P. 2013. Multi-view photometric stereo with spatially varying isotropic materials. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, Oregon, USA.

Zollhöfer, M., Nießner, M., Izadi, S., Rehmann, C., Zach, C., Fisher, M. & Wu, C. 2014. Real-time non-rigid reconstruction using an RGB-D Camera. ACM Transactions on Graphics, **33** (4): 156.

Submitted: 11/06/2016
Revised : 24/10/2016
Accepted : 29/11/2016